



Anderson, A. E. (2019). "We tried to, but life gets in the way": The Value of Cognitive Interviewing for Testing a Questionnaire on Antibiotic Consumption Behaviours. Unpublished. <https://doi.org/10.31235/osf.io/zyq2t>

Early version, also known as pre-print

License (if available):  
CC BY

Link to published version (if available):  
[10.31235/osf.io/zyq2t](https://doi.org/10.31235/osf.io/zyq2t)

[Link to publication record in Explore Bristol Research](#)  
PDF-document

This is the submitted manuscript (SM). It first appeared online via ScoArXiv at <https://osf.io/preprints/socarxiv/zyq2t/>. Please refer to any applicable terms of use of the author.

## University of Bristol - Explore Bristol Research

### General rights

This document is made available in accordance with publisher policies. Please cite only the published version using the reference above. Full terms of use are available:  
<http://www.bristol.ac.uk/pure/about/ebr-terms>

# **“We tried to, but life gets in the way”: The Value of Cognitive Interviewing for Testing a Questionnaire on Antibiotic Consumption Behaviours**

Alistair Anderson  
a.anderson@bristol.ac.uk

PhD Candidate  
School of Geographical Sciences  
University of Bristol (UK)

## **Abstract:**

The public’s perceptions and behaviours regarding antibiotics and antibiotic stewardship are often measured using survey questionnaires. Previous qualitative testing of questionnaire instruments that include questions about antibiotics or drug-resistant infections has suggested that questions on this topic are subject to measurement error for multiple reasons. This study used 13 cognitive interviews with pet owners to identify issues with a questionnaire instrument for a survey examining pet owners’ knowledge and behaviour around antibiotic use both personally and regarding their pets. Key findings from the study are that there are notable differences in quality of recollection between personal and pet-focused antibiotic use, and that socially desirable responding is only applied to certain behaviour questions. This article argues that cognitive interviews can provide substantial benefits as part of mixed-methods social research by supplementing dimension-reduced survey questionnaires with participants’ explanations, narratives, and experiences relating to the survey questions, with further benefits for qualitative theme generation.

## Introduction

Antibiotic resistance is a significant and growing challenge to healthcare systems. Whilst antibiotic resistance is a fundamentally biological phenomena, it has many social dimensions among its drivers and consequences. Community consumption of antibiotics is one such socially-patterned driver that has been previously examined through both qualitative and quantitative lenses. A common quantitative approach used for research into knowledge and behaviour around antibiotics is the survey questionnaire (for example Napolitano *et al.* (2013) and Vallin *et al.* (2016), in general community settings, Stallwood *et al.* (2019) with cat owners, and Fredericks *et al.* (2015) specifically regarding upper respiratory tract infections). To date there has been little qualitative empirical engagement with respondents' experiences answering survey questions about antibiotic use. Ensuring that questions are clear and understandable to respondents, that respondents are answering the questions intended by researchers, and understanding the variety of experience that is reduced by questionnaire categories are all areas of interest for survey research into public antibiotic stewardship. This article presents the qualitative testing of a survey questionnaire instrument which examines pet owners' knowledge and behaviour around antibiotic use personally and in administration to their pets, and in doing so addresses these areas of interest in this context.

There are some apparent issues for survey questions about antibiotic use, evidenced by the removal of questions on drug-resistant infections following the testing conducted for the 2018 global wave of the Wellcome Monitor. The Monitor's 'Questionnaire Development Report' noted that in multi-country cognitive testing both the notion of drug-resistant infections and the term 'antibiotic' were subject to confusion and differences in interpretation dependent on respondents' education levels or socio-economic status, leading to the decision to remove the questions (Gallup, 2018). This highlights that the findings of questionnaire-testing relating antibiotics are not necessarily generalisable across geographic contexts, or between regions with markedly differing levels of development. Questionnaire instruments should therefore be tested where possible in different settings, and the findings of the testing disseminated clearly and systematically.

One qualitative approach to evaluating the performance of survey questions is cognitive interviewing. Cognitive interviews are predominantly concerned with the process of data generation, and aim to provide information about how respondents arrive at their answers and uncover difficulties or ambiguities that are faced by survey respondents on the

way to their responses through verbal think-aloud techniques as well as interview probes (Campanelli, 2008). Cognitive interviews examine whether the questions fulfil their intended purposes and test the assumption that the meaning of the question intended by the survey author is consistent with respondents' interpretation of the question (Willis & Artino Jr., 2013; Dietrich & Ehrlenspiel, 2010). These techniques provide several benefits over quantitative pilot tests, for example through evidencing whether the intent of a closed question is being correctly understood by participants<sup>1</sup> (Campanelli, 2008; Buers *et al.*, 2014) or capturing the variability of conceptualisations of terms such as "health" that may be present in survey questions (Boeije & Willis, 2013). Whilst cognitive interviewing may not permit researchers to make assumptions about the true number of problems in a questionnaire, nor the problems that may arise with untested specific groups in the general population, they can produce data even at small sample sizes that can greatly improve confidence in a survey instrument's quality of data collection.

This cognitive interviewing study was part of a broader mixed-methods research project examining pet owners' knowledge, perceptions, and behaviour regarding antibiotic consumption and antimicrobial resistance and the role of online health information in these areas. The survey questionnaire instrument for the research mobilised some original questions, but also adapted questions from previous studies where testing information was not available in detail. To evaluate respondents' engagement with the questions the questionnaire instrument was tested with two rounds of cognitive interviews, the findings of which are presented in this article.

## **Previous Cognitive Interviewing Studies Covering Antibiotic-Related Questions**

Some previous cognitive interviewing studies have included antibiotic-related survey questions. These have however generally been limited to questions that mention antibiotics as context or example rather than as main study focus; for example as a term to be defined (Lapka *et al.*, 2008), as part of a study on misinterpretations of drug label instructions (Wolf *et al.*, 2007), as part of blood donor screening questionnaires (Beatty, 2002; Willson *et al.*, 2013), or the epidemiology of drug-resistant infections (Macario *et al.*, 2010).

A common finding relating to comprehension is that some respondents will exhibit confusion over what an antibiotic is and what they should be used for (Beatty, 2002; Wolf *et*

---

<sup>1</sup> In this article, participants and respondents are distinguished such that 'participants' refers to interviewees whilst 'respondents' refers to potential survey respondents.

*al.*, 2007; Lapka *et al.*, 2008; Macario *et al.*, 2010; Willson *et al.*, 2013). A potential consequence of this is that respondents do not report having taken an antibiotic when they have done so because they are not aware that a certain medication was an antibiotic. The reverse is also possible, with respondents possibly reporting having taken an antibiotic when in fact they have actually taken cold medicine or other over-the-counter remedy. The extent of this finding is dependent on the sample and the specific question being posed, with some studies also finding participants very able to articulate specifics relating to antibiotics, such as their purpose or specific instances in which they had been taken (Beatty, 2002).

A second common set of findings relate to information retrieval. Issues relating to participants' recollection include specifics of prescriptions (Macario *et al.*, 2010), and again, whether what had been taken in a specific instance was actually an antibiotic (Beatty, 2002; Macario *et al.*, 2010; Willson *et al.*, 2013). These findings demonstrate the variance in the memorability of antibiotic consumption, with implications for the quality of data collection regarding the reason for consumption and the method of consumption (for example, when measuring prescription adherence). These varied findings may be expected given the range of situations in which antibiotics can be consumed, from mild or preventative cases through to severe infections.

## **Aims and Justification of the Study**

There were two aims in this cognitive interviewing study. The main aim was to identify problems in individual items that would increase measurement error or respondent burden if left uncorrected. The second aim of the project was to test the questionnaire instrument as a whole and attend to contextual effects such as leading questions. Knowing the extent of any leading effects was important due to the use of duplicated questions in the first and third sections of the questionnaire covering human and pet health respectively. The correction of issues uncovered by this qualitative pretesting was the central objective of this stage of the overall research project.

The questionnaire instrument drew in part upon two previous studies' questionnaires that used differing modes of data collection. The Wellcome Trust Monitor Wave 3 (Wellcome Trust & Ipsos Mori, 2016a) used a face-to-face mode of administration, and the European Food Safety Authority's (EFSA) "Perceptions on the human health impact of antimicrobial resistance (AMR) and antibiotic use in animals across the EU" study (ICF, 2017) used an online mode of administration for its survey element. Understanding how

effective these questions – and adapted versions of these questions – were in this new questionnaire context, and whether they translated effectively across modes and questionnaires, was important for ensuring reliability and validity of questionnaire items and reducing measurement error in the survey study.

This study contributes to social research in the area of antibiotic use by presenting and discussing qualitative research into the construction of a questionnaire on antibiotic use and knowledge, with implications for future social measurements in this area. The study also presents some findings regarding pet owners' antibiotic-related knowledge and behaviour, which is an area that has only recently begun to receive substantive attention in either qualitative (Smith *et al.*, 2018; Redding & Cole, 2019) or quantitative (Stallwood *et al.*, 2019) literature.

## **Key findings**

Key findings from the cognitive interviews relate to differences in the quality of behaviour recollection between personal and pet-focused antibiotic use, selective demonstration of social desirability in responding to behaviour questions, and issues with the adaptation of questions from face-to-face mode surveys to a self-complete mode. Through the analysis and discussion of these findings, this article provides a contribution to future antibiotic-related survey questionnaire design by highlighting the benefits of qualitative testing for the questionnaire and the broader research project within which it was mobilised. These benefits include the suggestion of areas where the survey approach itself may compare unfavourably to qualitative approaches, the supplementing of closed-question responses with the interview sample's experiences and narratives around the situation the question was capturing, and the suggestion of further areas for research prompted by the cognitive interviewing data.

## **Research Design**

### **Data Collection Approach**

The study's ethics application was signed off through University of Bristol Research and Enterprise Development (RED) (ID: 60441) in January 2018. All interviews were conducted face-to-face, with all participants receiving an information sheet upon indication of

interest and both written and verbal briefings at the point of interview. Consent was provided in written form prior to interview commencement.

The interviews for this study were conducted in two waves, with variation in sampling approach between waves. All but one of the interviews were recorded and transcribed with participants' permission, and the questionnaire and interview protocol were adjusted between waves. By using two waves of interviews the researcher (acting as both interviewer and analyst) was able to iteratively test and improve the questionnaire instrument. The iterative testing was not open-ended in this study due to both time and budget constraints. This also meant that saturation was not a goal of the study, but rather the aim was to identify and resolve the most serious issues with the questionnaire before its deployment.

The interviews employed a think-aloud protocol followed by retrospective probing. During the think-aloud exercise, interview participants read each question aloud and actively verbalised their thoughts on their route to an answer that fitted the options provided. Think-aloud was used because the approach provides participant-initiated data at the point of answering the question (D'Ardenne, 2015), and as the survey was eventually to be self-administered with no interviewer present this data was considered more valuable than that which might be obtained through more disruptive concurrent probing. Retrospective probes were employed after the participant had completed the questionnaire. This is in contrast to concurrent probing, which entails the interrupting of the participant's flow through the questionnaire in order for the researcher to ask probes. Additionally, concurrent probing was avoided in order to facilitate the assessment of continuity in the questionnaire between the duplicated health sections.

The retrospective probes were developed to be employed where the think-aloud exercise did not cover an area of interest (for example, concepts that went undefined by participants). Specific probes included category-selection probes, for example clarifying the interpretation of response categories that were unspecific quantities of time (for example, 'once per week' compared to 'several times per week' or 'multiple times per month'). The same kind of probe was used for Likert-type items, differentiating between agreeing or disagreeing and 'strongly' agreeing or disagreeing. Specific concepts such as 'trustworthiness' or internet-based 'information sources' were probed to examine consistency of definitions where verbal reports had not explicitly tied the concepts to definitions that manifested in response judgements.

## Participant Selection

Thirteen participants were recruited in total. In the first wave, six postgraduate student pet-owners were recruited through convenience and snowball sampling. The use of participants with higher levels of education in this round follows Ackerman & Blair's (2006) proposition that it is more productive (where problem identification is the aim) to over-recruit respondents with above average education. In Ackerman and Blair's (2006) study, respondents with higher education levels yielded higher numbers of problems per interview because they spent more time either thinking about or discussing each question, recognising potential problems as well as encountering actual problems. In this first sample, highest education levels were distributed evenly between undergraduate degrees (n=3) and postgraduate degrees (n=3). The second round of interviews recruited seven pet-owners in the local area through a social-media based convenience sample. Highest education levels in this sample included GCSE level (n=1), A-level (n=2), NVQ Level 4 (n=1), undergraduate degree (n=2) and postgraduate degree (n=1). The first sample was intended to identify a higher frequency of issues, the second sample was intended to test the questionnaire with the more general community of pet-owners that were more reflective of survey's eventual respondents. Interviews were conducted at locations chosen by the participants, including participants' homes, cafés/public houses, and the interviewer's home.

Table 1 lists descriptive statistics for other key demographics. The first sample were generally younger than the second, with both samples slightly skewed towards younger age. The combined sample contained more female than male participants. Cat-only owners were most common in the sample, followed by participants owning a mix of pets (an exclusive category; for example, an owner of both cats and dogs would be counted once as 'Mixed'). As the objective of the study was not to collect representative survey data, the demographic characteristics of the sample are not vital to the validity of the study itself. The demographic data do however demonstrate that the cognitive interviews covered a diverse group with regards to the key demographic variables of the questionnaire. Multiple configurations of pet ownership were represented in the cognitive interviewing study, including sole smaller pets such as rodents and mixed pet-ownership (for example, dogs and chickens, or cats and guinea pigs).

		Wave 1	Wave 2	Overall



*Table 1. Sample demographic characteristics*

<b>Age</b>	<b>Range</b>	<b>22-38</b>	<b>21-53</b>	<b>21-53</b>
	<b>Median</b>	23	31	26
	<b>Mean</b>	26	33	30
<b>Gender</b>	<b>Male</b>	4	1	5
	<b>Female</b>	2	6	8
<b>Pets</b>	<b>Only Dogs</b>	1	1	2
	<b>Only Cats</b>	3	3	6
	<b>Only Small Animal</b>	1	0	1
	<b>Mixed</b>	1	3	4

## Data Analysis

Transcripts were coded initially by the individual issues that arose, with twenty-one individual issues relating to antibiotic use or internet questions identified across the two rounds of interviews. This enabled an appreciation of the range and recurrence of issues. The individual codes, combined with specific comments from participants, were the eventual basis for decisions over how a question/response might be altered. These individual issues and their codes were divided into three overarching categories. Firstly, those issues that were both recognised and verbalised by participants and hindered participants' responding in some way, for example through confusion over a question's wording or an inability to make a clear judgement within the categories provided. Secondly, those issues that were not recognised but were verbalised by participants. These were issues that mainly provided a post-survey analytic problem such as participants missing a question's request for the most recent instance and verbalising their most memorable instance of antibiotic use, or telephone consultations with doctors being selected as 'Other' when the aim of the question was to deduce whether there had been a consultation of any type. Finally, there were potential issues that were identified and verbalised but that did not directly affect the participants' response where participants spontaneously suggested scenarios in which a concept or phrase might be a challenge for a hypothetical respondent.

These overarching categories gave an overview of the mechanics of the questionnaire for survey respondents – for example where interview participants were actively struggling to respond to the questions or were passively misinterpreting the aims of the questions. These categories, combined with the initial coding, provided a basis for deciding whether a question/response should be altered. For example, recurrent misinterpretations of a concept being recognised and verbalised by respondents would certainly need resolving, whereas a hypothetical syntax problem that was only verbalised by the participant that ‘spotted’ it would be less likely to require alteration. The breakdown of issues within these three categories is presented in Table 2, with issues ordered by decreasing frequency in the first round.

The total number of issues decreased between rounds. This could be attributable to the improvement of the questionnaire instrument, though it could also be reflective of Ackerman and Blair’s (2006) suggestion that a sample of higher-educated participants yield higher numbers of issues. The most common issue was with recollection, as participants most often consciously struggled to recall the details of the last time they or their pets had consumed an antibiotic because of the time elapsed between the event and the question being posed. The second most common issue was the potential for a participant to select multiple answer options where only one was requested – an urgent issue to be addressed before dissemination of the survey. Leading effects between questions did not appear to be a significant problem for the questionnaire instrument, with only one instance of an answer being deduced from a previous question. The volume of issues identified suggests that the exercise was a useful one in the process of the questionnaire’s development.

Table 2 Frequency of issues per round

Issue Type	Issue	Round 1 Frequency	Round 2 Frequency	Total Frequency
<b>Recognised by Participant</b>	Recollection	10	4	14
	Multiple Potential Response Options	3	2	5
	Aim of Question Unclear	1	1	2
	Inflation Due to Employment-Related Information Searching	1	1	2
	Misreading	1	0	1
	Insufficient Response Options	0	1	1
	More Examples Requested	0	1	1
<b>Not Recognised by Participant</b>	Responded with Household Member's Actions	4	0	4
	Animals Included in Human Response	0	3	3
	Recollection	1	1	2
	Desired Information Reported as 'Other'	0	1	1
	Misreading	1	0	1
	“Maybe” Reported as True	1	0	1
	“Don’t Know” Reported as True	0	1	1
	Deduction from Previous Question	1	0	1
	Non-Average Level of Use	1	0	1
	“Yes” Ticked as “N/A”	0	1	1
<b>Hypothetical Identified by Participant</b>	Category Overlap	1	0	1
	Definition of 'Healthcare Professional'	1	0	1
	Syntax Issue	0	1	1
<b>Total Issues</b>		27	18	45

# Findings

The findings of the cognitive interviews are presented by questionnaire section (Internet-related questions, followed by antibiotic-related questions). Issues are organised thematically based on the areas in which they arose. The questionnaire itself was structured with the human-related section coming first, followed by a demographic information section, and finished with the pet-related section which was a mirror of the wordings of the human-related section.

## Internet-Related Questions

### Framing ‘Information Sources’

In a question referring to the trustworthiness of internet-based information sources relative to doctors, there was some inconsistency in the framing of “information sources that you use on the Internet”. In the first round, four participants referred to the internet as a general resource, whilst two referred specifically to the National Health Service’s (NHS) website relative to their GP. The words ‘in general’ were added to the start of the questions, which resolved this issue in the second round of interviews and increased consistency across respondents in terms of how they referred to the internet as a resource. Respondents compared a variety of information sources to their veterinarians and lacked consensus on an ‘official’ online information source for companion animals. It is important to not underestimate the prevalence of use of ‘official’ online information sources such as the NHS website when examining general use of online sources of health information in this format.

### Consideration of ‘Trustworthiness’

The relative trustworthiness of websites, doctors, and veterinarians was considered in some depth by most participants. Some participants discussed the different diagnostic approaches between internet-based self-diagnosis and the practices of healthcare professionals. Healthcare professionals were often perceived as having a more holistic view of symptoms and potential diagnoses, with GPs for example “*know[ing] a little of a lot of things*”, or whilst websites provide “*reasonably true*” information a GP has the ability to point out the rarity of suggested diagnoses and “*do those eliminations for you*” due to their “*experience of making those judgements*”. Animal health was more directly problematic for participants, as there was general consensus that the lack of an equivalent to the NHS website in the context of animal health made the assessment of source quality more difficult. For

some, making direct comparison with the NHS, this meant the lack of a “*sounding board*” for their pet’s health, whilst others highlighted a lack of “*developed trust*” in specific Internet presences. There were also substantive differences between participants in their comparison of doctors and veterinarians to respective online information sources. For one participant, whilst there was a lack of developed trust in a specific animal health website vets also had “*less authority with me than doctors would. With the client relationship, they have less authority*”. Conversely, for another participant this perception was reversed: “*I can feel my own pain and I can’t feel the animals’ pain, so I have to take [the veterinarian’s] word for it. Whereas with me, it’s like, I’ll do what I want*”. As such reflections are often reduced in survey questionnaires to closed response categories, such as Likert-type items, supplementing the survey data with a set of qualitative interpretations of key concepts in the questionnaire can provide valuable insight into the diversity of intentions and experiences behind the more dimensionally-reduced survey question responses. Furthermore, in a mixed-methods project such as the one from which these findings are drawn, these data can prompt further exploration through less structured qualitative interviews or focus groups alongside the improvement of the survey questionnaire instrument.

### **Frequency of Internet Use**

Another area that led to inconsistencies in judgement rationales was the frequency of Internet use. For the question “How often do you use the Internet to search for health information relating to humans” (or, in its mirrored version, “...for animals”), one participant in the first round verbalised different times at which they had used the internet more frequently with regards to their pet without explicitly averaging their frequency of use for their answer, commenting that they used the internet for pet health “*when my cat was sick, every day when she was sick*”. Additionally, one participant in each round referred to their main working activity as involving health either in the context of research or direct provision of care. In both cases, this boosted the frequency with which they reported using the internet to search for health information relating to humans as it was unclear as to whether the question was “*about work, or about me, or people that I know*”. “On average” was inserted at the start of the question and was picked up on by second round participants, with some participants verbalising clustered behaviour as with the first round but explicitly translating this into a perceived average frequency of behaviour. Generally the internet was only being used when there was a specific cause for concern or where pre-existing conditions were involved, and in reporting their judgement of an average level of use respondents were

conveying how habitually they would turn to the internet as a source of information on health topics. This example emphasises the value of qualitative testing of survey questions through cognitive interviews, as the reliability of this question in its untested form was negatively impacted by specific circumstances that were skewing some participants' comprehension of the question.

## **Antibiotic-Related Questions**

### **Memorability and Recollection**

In the human-focused questions, participants that had difficulty recalling the relevant information generally did so because the last time antibiotics had been taken was either a long time ago, for something that was not memorable, or both. Three participants in each round presented some difficulty with recollection for the initial question "Please think back to the last time *you* took antibiotics. Where did you get those antibiotics from?". One participant in the second round who had presented a recall issue with this question also presented a recall issue with the next question, "How did you take these antibiotics?". This recall issue did not prevent an answer from being rationalised by the participant, however. In this instance, the participant verbalised "*I probably took them until I felt better. I'm not a fan of taking medication*" and chose the option "Taken until you felt better" on the questionnaire. This suggests that some respondents will extrapolate from their underlying beliefs about medication when recall presents a barrier to response, even when these beliefs may be socially undesirable. Both the human- and pet-focused strands of the subsequent question "What did you do with any leftover antibiotics?" presented recall issues for participants, once on the human side after probing and three times for the pet side during think-alouds. A consequence of one recall issue was one participant checking two responses after the following verbalisation: "*Since I'm not sure – if there were leftovers and they didn't get used, we would have kept them, otherwise there were none left over*". A "don't know" option was not provided in this question so as to maximise the provision of substantive responses, and the lack of such an option generally prompted all other participants to select a substantive response category to the best of their knowledge. However, following this last example, "Please check one option" was appended to the question between rounds to emphasise the selection of a single response, with no double-selections considered by second round participants.

Conversely, some participants reported high levels of clarity in their recollections of taking antibiotics themselves or administering them to their pets. Two participants in the first round and four in the second reported that it was easy to remember the last time that they had taken antibiotics themselves either because it was recent, memorable, or both. Memorable reasons were not always attributable to the antibiotics themselves – whilst one participant did recall severe side-effects during a long course of antibiotics, another recounted a serious leg break that later became infected and, almost tangentially in their verbalisation, required antibiotics. Recall issues were less prevalent in the pet-focused iteration of “Where did you get those antibiotics from?” (four pet-related recall issues, with six human-related recall issues) but were more prevalent for “What did you do with any leftover antibiotics?” than for the human-focused versions (three pet-related recall issues, with one human-related recall issue). Multiple participants, particularly in the second round, noted that it was easy to remember the most recent instances of antibiotic use for their pets as they were particularly stressful or upsetting. All participants noted that it was easy to pick a response to the pet-focused acquisition question either because they would only ever get pet medication via a prescription from a vet, because the pet’s condition was a “*vivid memory*”, or because the mode of acquisition (for example, from friends or family) meant that they avoided a stressful trip to the vets with a reactive pet.

In general, there were more recall-related issues presented by participants with regards to their own antibiotic consumption than with regards to their pets. This is likely down to the higher number of factors involved in managing a pet’s health with medication, including transport of the pet, veterinary bills, and actually administering the antibiotics. Beyond questionnaire improvement, these findings may be suggestive of substantively different dynamics in antibiotic consumption behaviours by pet owners in terms of their own personal consumption and their administration of antibiotics to their pets. This is an area that is part of the broader research agenda within which this questionnaire and its testing reside.

### **Translating Response Categories Between Survey Modes**

The section-opening question “Please think back to the last time you took antibiotics. Where did you get those antibiotics from?”, which was adapted from the Wellcome Monitor Wave 3 questionnaire (Wellcome Trust & Ipsos Mori, 2016b), precipitated multiple category-overlap issues. These included one participant who had obtained antibiotics abroad seeing overlaps for the category “from abroad” with other response categories such as “prescribe

after face-to-face with a healthcare professional”, two participants translating telephone consultations into the “Other” response category, and one participant selecting both “Prescribed after face-to-face with a veterinarian” and “Online service with a prescription” having received the prescription from the veterinarian and subsequently bought the antibiotics online. In each of these cases the participants’ scenarios had categories in which the researcher intended them to fit, but the participants interpreted multiple response categories as potentially appropriate.

In response to these issues, changes were made between rounds and following the second round. Between rounds the phrase “health professional” in the human-focused question was changed to “general practitioner or nurse”, and “from abroad” was changed to “Other” to act as a catch-all for the various possible configurations of antibiotic consumption abroad that were not included in current categories. Following the second round, the question wording itself was changed with “last” becoming “most recent” with underscoring, and the response category “prescribed after face-to-face with a general practitioner” again being altered to “In person following prescription from oral consultation with a general practitioner or nurse”. Whilst this is a more verbose formation, it was considered acceptable because it was necessary to delimit the category sufficiently from “online service with a prescription” as well as provide an option for phone consultations. Whether a respondent’s consultation was in person or over the phone was not analytically important for the survey, but the general avenue of antibiotic acquisition was. With regards to the pet-focused acquisition question, the question and responses were altered to mirror the human question which addressed the category overlap issue presented for this question. The key issue in these examples was the translation of response categories from a survey administered through face-to-face interviews into a self-administered mode of data collection. The cognitive interviews here served to bridge the gap left by the lack of an interpreting interviewer by highlighting specific required changes and consequently improving the validity of the question as a measurement tool.

### **Social Desirability when “Life Gets in the Way”**

When responding to the questions about how antibiotics had been consumed/administered, two participants verbalised one response and consciously selected a different response. Specifically, these participants selected “Taken as prescribed and at the correct times” when they verbalised that they had not done so. For example, one participant responding to the human question verbalised that the antibiotics were taken as “*probably a*



*mixture of the second and third options [...] but generally I would say taken as prescribed at the correct times*". This issue recurred in the mirrored pet question, both with the same participant who reflected "*Again, probably a mixture of correct times and not correct times*" and another participant who recalled "*usually at the correct times, but not always.*" When prompted to settle on one category or the other, this second participant answered that they would select "[...] *given as prescribed at the correct times. We tried to, but life gets in the way.*" These responses suggest that some respondents to the questionnaire may simply report that they had taken (or given) antibiotics at the correct times when there were instances where they had not done so. This challenges the reliability of the question as a measurement tool, because the implication is that the level of antibiotic consumption with incorrect timings will be underestimated while correct timings are overestimated. As the main aim of the item(s) in question was to assess the difference between respondents that stop taking antibiotic courses when they feel better as opposed to those that finish their prescribed course as instructed, distinguishing between following a prescribed course at the correct or incorrect timings was deemed not to be a useful distinction to require given this evidence. Qualitative research by Hawkins *et al.* (2008) suggests that individuals that *intend* to take the full course of antibiotics take their medicines at 'mostly' the correct times and regret missing specific doses, which is reflected in this behaviour with the questionnaire. Consequently, these two response options were reduced to a single 'taken as prescribed' option in order to reduce respondent burden. This example demonstrates that such specific response categories should be interpreted with caution when behaviour is examined through dimension-reducing tools such as survey questionnaires. Qualitative approaches, such as interviews, may be more reliable for engaging with antibiotic consumption behaviour where this level of detail is desired.

### **Leading Effects**

There was some evidence of a leading effect between knowledge questions. All participants responded that it was true that unnecessary use of antibiotics in humans could lead to antibiotics becoming ineffective to treat humans. However, in answering a similarly worded item relating to use in animals affecting antibiotic efficacy for humans (which followed directly from the human use question) one participant moved from a "Don't Know" – which is a substantive response for this question – response to "True" based directly on deduction from the previous (human use) question. This question and the previous item were switched in position between rounds, with no leading effect observed in the second round.

## Diversity of Experience

A second-round respondent that had made two selections on the question about where they had acquired their antibiotics due to obtaining them via a telephone consultation also misread the question and recalled multiple events rather than only the most recent. This contributed to their rationale for selecting multiple response categories, as they were reporting multiple events rather than one. The same participant that missed the recency aspect of this question about acquisition also missed it in their calculation of an answer on the human branch of the question “What did you do with any leftover antibiotics?”, verbalising about both “*a time I was prescribed antibiotics, a long course where I was allowed to stop them when I wanted to*” and “*a regular occasion*”. Being in the second round, they did however make an explicit note of the instruction to check one box and selected their response as if for a “*regular*” occasion. This participant reported having a long-term health condition and having to take courses of antibiotics multiple times per year, so the high frequency may have had the same effect as distant recall in making a specific recent instance difficult to bring into focus. This highlights the importance of qualitatively testing a questionnaire as this can ensure that the survey questions are both accessible and reliable measurements across different respondent backgrounds, and in this case medical backgrounds.

## The Role of Examples

One participant suggested, after probing, that more examples would assist with the answering of “In humans, what conditions do you think can be effectively treated by antibiotics? (Tick all that you think apply)”, because the participant was “*not medical at all*” as a person but could recall the different infections that they had had treated with antibiotics previously. Conversely, the examples provided in the responses successfully triggered another participant into ticking “bacterial infections” as opposed to just “viral infections”. Whilst an exhaustive list could be provided (or indeed, none at all), the question was concerned with respondents’ generalisations of types of infections related to antibiotic use as opposed to specific infections and the use of a small number of examples was intended to make the question more accessible with some common infection examples without increasing respondent burden rather than provide an explicit structure to a participant’s recollection. This example illustrates the value of qualitative testing in examining the role of provided examples in respondents’ experiences of the survey and clarifying the extent to which the examples are help or hindrance.

## Discussion

The use of a qualitative approach – cognitive interviewing – to test the survey questionnaire within this mixed-methods research project added considerable value to the research in several ways. Firstly, specific challenges to the reliability and validity of questions as measurement tools were able to be identified and corrected with targeted adjustments. Secondly, the narratives that participants provided whilst answering the questions add value to the later interpretation of quantitative analyses that, whilst providing a level of generalisability and the benefits of statistical modelling, involve substantial dimension reduction that may limit consideration of the nature and diversity of research participants' experiences with the phenomena of interest. Finally, the qualitative data generated by this method is valuable for the development of other areas of the research project such as further qualitative investigations and theory development.

In this questionnaire, the first group of issues arising from questions on the use of the internet referred to the implications for analysis of how participants were interpreting the specificity of the internet use they were being asked about. More generically, there were issues relating to time frames that in some cases impeded response judgement or presented caveats for the later analysis of survey data. These examples demonstrate the value of qualitative testing of this questionnaire instrument, as challenges to both reliability and validity of the questions as measurement tools were raised and specific alterations could be effected by the researcher to address them.

A substantial amount of qualitative data beyond solely the testing of the questions was collected as participants discussed their understandings of concepts such as 'trustworthiness' and how they related to health information across human and animal healthcare. These data highlighted substantive differences in the consideration of human and veterinary health professionals and the consideration of online health information sources by participants, and provide qualifications with regards to the rationalisation of different responses to identical questions between each domain of health. Furthermore, the differences in issues related to recollection between personal antibiotic consumption and administration of antibiotics to pets highlight another aspect of the different ways participants related to the domains of human and veterinary medicine. Whilst the quality of this difference is a substantive area of interest for the wider mixed-methods project, the effect of this difference on the survey questionnaire's capability as a measurement tool is an important finding from the cognitive

interviews for the analysis of the survey itself in terms of the potential difference in measurement error between respondents' answers for their personal behaviour and their behaviour in administering medication to their pets.

The demonstration of recollection issues by participants with regards to their consumption of antibiotics were generally commensurate with previous examples of cognitive interviews that covered antibiotic consumption. Reasons for difficulty of recall often mirrored those for ease of recollection – for some participants the last time that they had taken antibiotics was a long time ago and/or not for anything memorable, whilst for others it was recent and/or vivid. With regards to the pet-related side of the questionnaire, those respondents that had given their pets antibiotics tended to remember more clearly what they had been given for compared to their personal consumption, especially in some cases where the pet-related event was a particularly stressful or upsetting occurrence. As antibiotics may be taken in a range of scenarios, from mild or preventative cases through to severe infections, the variation in recall issues among participants may be expected though this would not necessarily be clearly reflected in the survey data alone.

In another example, whilst recall of specific instances of antibiotic use in pets was less of a problem than in personal use, recall issues were more pronounced for action taken regarding pet-related leftover antibiotics, with multiple participants verbalising such issues spontaneously during their think-aloud compared to a single participant bringing it up with regards to personal use only after being probed. This suggests that recall issues for the same instance of antibiotic consumption manifest differently in questionnaire responses for different aspects of the procedure of acquiring, taking, and keeping/disposing of antibiotics. This too would not be an issue that could be clearly uncovered through a simple pilot test of the questionnaire nor would it be apparent in survey data, and with the previous examples is further suggestive that the survey measurement of antibiotic-related behaviour in personal and pet-related contexts have differing amounts of measurement error even with identical questions.

As behaviour questions increased in specificity, there was evidence that for some questions their validity was reduced due to socially desirable responding. There were cases where participants' verbally recalled actions were different to those they reported in their mock questionnaire with regards to the timing of antibiotic consumption. The challenge to the validity of the split between taking or giving antibiotics with correct or incorrect timings was

significant enough to require that the response categories be combined into a single category simply measuring whether antibiotics were taken as prescribed or not. If the distinction between the correct and incorrect timing of consumption for respondents is an area of interest, less dimensionally-reduced forms of data may better serve researchers. Previous examples of qualitative research have, for example, elaborated on this area in detail (Hawkings *et al.*, 2008).

## **Strengths and Limitations**

A strength of this study is that despite the small sample – itself not unusual or inhibitory for cognitive interviewing studies (Collins & Gray, 2015; Beatty & Willis, 2007; Boeije & Willis, 2013) – the study included participants at multiple stages of life, with multiple levels of education, and specifically for this study’s overall purpose, multiple configurations of pet ownership. A second strength is that this study examined multiple specific facets of antibiotic use and knowledge in the questionnaire – including acquisition, behaviour, knowledge of antibiotics’ function, and knowledge of antibiotic resistance – in the context of both personal use and pet-orientated use.

With a larger sample and further rounds of refinement, more issues with the questionnaire would certainly have been uncovered. Blair & Conrad (2011) have demonstrated that significant issues can still be found even after 70 cognitive interviews – though with diminishing returns as the sample size grows. Due to limitations of budget and time however, saturation was not the aim of this study. Further studies could supplement the findings of this study by testing antibiotic-focussed questionnaires in a variety of other healthcare settings and scenarios.

## **Conclusion**

This study used a qualitative method to test a questionnaire instrument for a survey covering pet owners’ knowledge of antibiotics, antibiotic use behaviour, and use of the internet for health information. The main objective was to uncover problems with questions that would affect respondents’ abilities to respond or would increase measurement error in the survey. Qualitative testing of questionnaires can provide complementary value to survey analysis by elaborating on the experiences of a subset of participants in direct relation to the questions and their response categories. This is important because while survey research has

significant strengths in terms of generalisability and the quantification and controlling of associations between variables in analyses, these strengths come at the cost of dimensionally reduced data. Such reduced data can mask sources of measurement error in questionnaires if not tested thoroughly, exemplified by the dissonance between verbalisations and actual questionnaire responses regarding the timing of antibiotic consumption discussed in this article. Moreover, the cognitive interviews can serve as a bridge between methods in mixed methods projects, connecting the more rigid form of data collection in the survey questionnaire to the more fluid and often spontaneous data collection of interviews and focus groups. An example of this is the discussion of trustworthiness by cognitive interview participants in the specific context of the questionnaire questions, an area that has since been examined in greater depth during semi-structured qualitative interviews in another part of the research project.

In general, the most prevalent issue for questions about previous antibiotic use in this study was recall. This was an issue identified in prior cognitive interviewing literature involving questions about participants' previous behaviour with antibiotics. This prevalence did not mean that all respondents struggled to respond to questions regarding antibiotic consumption however. Participants in this study could generally recall at least some detail of an instance in which they had taken antibiotics, and where they could not do so explicitly they extrapolated from underlying habits and beliefs regardless of their social desirability. For studies aiming to measure or model such beliefs there do not appear to be serious measurement error problems with questions regarding antibiotic acquisition and consumption, based on the verbal reports of these participants. More specific behaviours are less reliably reported however, as demonstrated by the participants that verbalised that they had not managed to take antibiotics at the correct times for their prescription, but still selected the 'correct times' response category anyway.

Beyond questionnaire problem-finding, a key finding of the cognitive interviews was the substantive difference between participants' experiences with and rationalisations of responses to identical questions about their personal use of antibiotics and their administration of antibiotics to their pets. Firstly, this suggests that there may be different levels of measurement error in surveys that examine antibiotic-related behaviour in these two settings. Secondly, this suggests that this may be an area of substantive research interest.

Understanding trends in attitudes and behaviours in the context of community antibiotic consumption is an important part of the mitigation of antibiotic resistance. Cognitive interviewing can add significant value to research into the social patterning and individual rationalisations of behaviour with antibiotics both by improving the measurement potential of survey research, and by generating complementary qualitative data that can inform research conclusions and designs.

### **Acknowledgements:**

Thanks to Dr. Maria Fannin for providing critical and productive comments on previous drafts of this article.

### **Funding:**

This research was funded by an Economic and Social Research Council studentship grant (ES/J50015X/1).

### **Declaration of Conflicting Interests:**

The author declares no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

## References:

- Ackermann, A. & Blair, J. (2006). Efficient Respondent Selection for Cognitive Interviewing. *Proceedings of the Survey Research Methods Section of the American Statistical Association*, 3997-4004.
- Beatty, P. (2002). Cognitive Interview Evaluation of the Blood Donor History Screening Questionnaire: Results of a study conducted August-December, 2001. Retrieved from <https://www.aatb.org/sites/default/files/Blood-%20Cognitive%20Evaluation%20final%20report%20NCHS%20v2.pdf>.
- Beatty, P. C. & Willis, G. B. (2007). Research Synthesis: The Practice of Cognitive Interviewing. *Public Opinion Quarterly*, 71(2), 287-311.
- Boeije, H. & Willis, G. (2013). The Cognitive Interviewing Reporting Framework (CIRF): Towards the Harmonization of Cognitive Testing Reports. *Methodology*, 9(3), 87-95.
- Blair, J. & Conrad, F. G. (2011). Sample Size for Cognitive Interview Pretesting. *Public Opinion Quarterly*, 75(4), 636-658.
- Buers, C., Triemstra, M., Bloemendal, E., Zijnenberg, N. C., Hendriks, M. & Delnoij, D. M. J. (2014). The value of cognitive interviewing for optimizing a patient experience survey. *International Journal of Social Research Methodology*, 17(4), 325-340.
- Campanelli, P. (2008). Testing Survey Questions. In: E. D. de Leeuw, J. J. Hox, and D. A. Dillman eds. *International handbook of survey methodology* (pp. 176-200). New York & London: Taylor & Francis.
- Collins, D. & Gray, M. (2015). Sampling and Recruitment. In: D. Collins ed. *Cognitive Interviewing Practice* (pp. 80-100). London: SAGE Publications.
- Dietrich, H. & Ehrlenspiel, F. (2010). Cognitive Interviewing: A Qualitative Tool for Improving Questionnaires in Sport Science. *Measurement in Physical Education and Exercise Science*, 14, 51-60.
- D'Ardenne, J. (2015). Developing Interview Protocols. In: D. Collins ed. *Cognitive interviewing practice* (pp. 101-125). London: SAGE Publications.



- Fredericks, I., Hollingworth, S., Pudmenzky, A., Rossato, L., Syed, S. & Kairuz, T. (2015). Consumer knowledge and perceptions about antibiotics and upper respiratory tract infections in a community pharmacy. *Int J Clin Pharm*, 37, 1213-1221.
- Gallup. (2018). *Wellcome Global Monitor: Questionnaire Development Report*. Retrieved from [https://wellcome.ac.uk/sites/default/files/wellcome-global-monitor-questionnaire-development-report\\_0.pdf](https://wellcome.ac.uk/sites/default/files/wellcome-global-monitor-questionnaire-development-report_0.pdf).
- Hawkings, N. J., Butler, C. C. & Wood, F. (2008). Antibiotics in the community: A typology of user behaviours. *Patient Education and Counselling*, 73, 146-152.
- ICF. (2017). EU Insights – Perceptions on the human health impact of antimicrobial resistance (AMR) and antibiotics use in animals across the EU. *EFSA Supporting Publication 2017: EN-1183*.
- Lapka, C., Jupka, K., Wray, R. J., & Jacobsen, H. (2008). Applying cognitive response testing in message development and pre-testing. *Health Education Research*, 23(3), 467-476.
- Macario, E. Daum, R. S., Miller, L. G. & Eells, S. J. (2010). Using cognitive interviews to refine a household contacts survey on the epidemiology of community-associated methicillin resistant *Staphylococcus aureus*. *Journal of Infection Prevention*, 11(2), 44-48.
- Napolitano, F., Izzo, M. T., Giuseppe, G. D. & Angelillo, I. F. (2013). Public Knowledge, Attitudes, and Experience Regarding the Use of Antibiotics in Italy. *PLoS ONE*, 8(12), e84177.
- Redding, L. E. & Cole, S. D. (2019). Pet owners' knowledge of and attitudes toward the judicious use of antimicrobials for companion animals. *Journal of the American Veterinary Medical Association*, 254(5), 626-635.
- Smith, M., King, C., Davis, M., Dickson, A., Park, J., Smith, F., Currie, K. & Flowers, P. (2018). Pet owner and vet interactions: exploring the drivers of AMR. *Antimicrobial Resistance and Infection Control*, 7:46.
- Stallwood, J., Shirlow, A. & Hibbert, A. (2019). A UK-based survey of cat owners' perceptions and experiences of antibiotic usage. *Journal of Feline Medicine and Surgery*, 1-8.

- Vallin, M., Polyzoi, M., Marrone, G., Rosales-Klintz, S., Wisell, K. T. & Lundborg, C. S. (2016). Knowledge and Attitudes towards Antibiotic Use and Resistance - A Latent Class Analysis of a Swedish Population-Based Sample. *PLoS ONE*, 11(4), e0152160.
- Wellcome Trust, Ipsos Mori (2016a). *Wellcome Trust Monitor 3, 2015* [data collection]. UK Data Service. Serial Number: 7927. <http://doi.org/10.5255/UKDA-SN-7927-1>. Accessed 25/04/17.
- Wellcome Trust, Ipsos Mori (2016b). *Wellcome Trust Monitor Wave 3 Mainstage Questionnaire*. [Online]. Available at: [http://doc.ukdataservice.ac.uk/doc/7927/mrdoc/pdf/7927\\_wtm\\_w3\\_appendix\\_d.pdf](http://doc.ukdataservice.ac.uk/doc/7927/mrdoc/pdf/7927_wtm_w3_appendix_d.pdf). Accessed 25/04/17.
- Willis, G. & Artino Jr, A. R. (2013). What Do Our Respondents Think We're Asking? Using Cognitive Interviewing to Improve Medical Education Surveys. *Journal of Graduate Medical Education*, 5(3), 353-356.
- Willson, S., Miller, K. & Ryan, M. (2013). *Cognitive interviewing study findings of the uniform blood donor history questionnaire*. National Center for Health Statistics.
- Wolf, M. S., Davis, T. C., Shrank, W., Rapp, D. N., Bass, P. F., Connor, U. M., Clayman, M. & Parker, R. M. (2007). To err is human: Patient misinterpretations of prescription drug label instructions. *Patient Education and Counselling*, 67, 293-300.